On the analysis of grouped survival data using cumulative occurrence/exposure rates

Ian W. McKeague[1] and Mei-Jie Zhang[2]

Department of Statistics
Florida State University, Tallahassee, Florida 32306-3033

FSU Technical Report No. M-840
USARO Technical Report No. D-117
AFOSR Technical Report N0. 91-257

March, 1991

AMS 1980 subject classifications. 62G05, 62M09, 62J02.

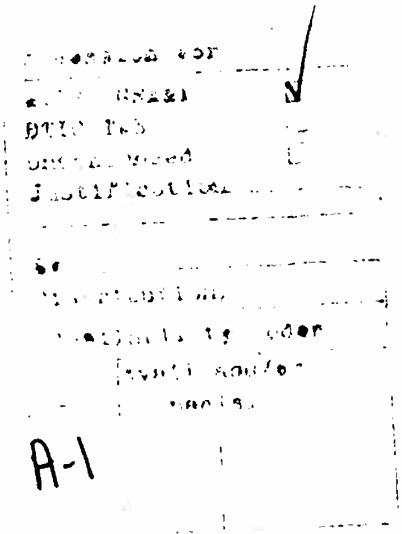
Key words and phrases. Counting processes, conditional hazard function, grouped survival data, martingale central limit theorem, proportional hazards model, nonlinear regression.

---

[1] Research supported by Army Research Office Grant DAAL03-90-G-0103.
[2] Research supported by the Air Force Office of Scientific Research under Grant AFOSR88-0040.

Abstract

One may estimate a conditional hazard function from grouped (and possibly censored) survival data by the time and covariate specific occurrence/exposure rate. Asymptotic results for cumulative versions of this estimator are developed, utilizing the general framework of counting processes. In particular, a grouped data based goodness-of-fit test for Cox's proportional hazard model is given. Various constraints on the asymptotic behavior of the widths of the calendar periods and covariate strata employed in grouping the data are needed to prove the results. Actual performance of the estimators and test statistics is evaluated by Monte Carlo methods.



A-1

1. Introduction

The purpose of this paper is to study grouped data based inference for Beran's (1981) general nonparametric hazard function model and Cox's (1972) proportional hazards model. In the general nonparametric model the conditional hazard function $\lambda(t|z)$ of the survival time of an individual with covariate z is given by

$$\lambda(t|z) = \alpha(t, z), \tag{1.1}$$

where α is an unknown function. In Cox's model, α is specified by

$$\alpha(t, z) = \lambda_0(t) \exp\{\beta_0' z\}, \tag{1.2}$$

where the covariate is p-dimensional, β_0 is a p-vector of unknown regression coefficients and λ_0 is an unknown baseline hazard function.

The usual approach to analyzing grouped survival data is to assume that α is piecewise constant over each grouping cell, see Hoem (1987). Then the likelihood function is proportional to the Poisson likelihood (Laird and Olivier, 1981) and the maximum likelihood estimator of α is the classical *occurrence/exposure* rate. Kalbfleisch and Prentice (1973), Holford (1976), and Prentice and Gloeckler (1978) have studied the maximum likelihood estimator of the regression parameters in Cox's model when the baseline hazard function is taken as a step function. Holford (1976) noted that this estimator is inconsistent unless the grouping becomes finer as the sample size increases.

It is important to know whether the convenience of analyzing grouped data from a given actuarial life table is overshadowed by biases that arise when the grouping is coarse. There exist many numerical studies comparing the grouped and continuous Cox model analyses for specific data sets, see the references in Hoem (1987, p. 137). All these studies have found that the two approaches give quite similar results. Breslow (1986), considering data on cancer mortality among Montana smelter workers, found that the estimated regression coefficients from the grouped data analysis were within one standard error of those from the continuous data analysis. Similar conclusions were reached by Selmer (1990) for data on mortality from coronary heart disease. Selmer obtained an extremely close agreement of likelihood ratio test statistics (used to test for differences between nested models) in the full Cox model and the Cox model with a piecewise constant baseline hazard function. However, it would be useful to have a theoretical underpinning for these empirical studies.

Theoretical results for *continuous* data are well developed; see Andersen and Gill (1982) (henceforth AG) for Cox's model, and McKeague and Utikal (1990a) (henceforth MU) for the general nonparametric model. Corresponding results for grouped data are available only in special cases. Aficionados of contingency table analysis might consult Friedman's (1982) paper on the Cox model, but his results are difficult to interpret in the survival analysis context. Pons and Turckheim

(1987) have used histogram sieve estimators for Cox's model (as have Borgan and Ramlau-Hansen (1985) and Karr (1987) for Aalen's multiplicative intensity model), but their approach applies to grouped data only when the covariate takes at most finitely many values and is non-time dependent. As far as we know, the general nonparametric model with grouped data has not been treated in the literature.

Our aim here is to show that the results of AG and MU have analogues in the grouped data case. Various conditions on the asymptotic behavior of the widths of the calendar periods and covariate strata used in grouping the data are needed for this. We also develop a grouped data version of MU's (1991) goodness-of-fit test for the Cox model. Estimators of the conditional cumulative hazard function $A(\cdot, z) = \int_0^{\cdot} \alpha(t, z)\, ds$ and the doubly cumulative hazard function $\mathcal{A}(\cdot, \cdot) = \int_0^{\cdot}\int_0^{\cdot} \alpha\, dt\, dx$ play an important role in this work. The goodness-of-fit test is based on a comparison of estimates of $\mathcal{A}$ under the general nonparametric model and the Cox model.

In Section 2 we formulate the general model (1.1) in the (by now standard) counting process setting and discuss the estimation of $A(\cdot, z)$ and $\mathcal{A}$. Our Cox model results, extending AG, are given in Section 3. The goodness-of-fit test for the Cox model is discussed in Section 4. Section 5 contains a simulation study and an application to real data. All proofs are contained in Section 6. For simplicity we restrict attention to the case of a one-dimensional covariate ($p = 1$) throughout the paper.

2. Fitting the general nonparametric model to grouped data

Let $\mathbf{N}(t) = (N_1(t), \ldots, N_n(t))'$, $t \in [0, 1]$ be a multivariate counting process with respect to a right continuous filtration $\mathcal{F}_t^{(n)}$, where $N_i(t)$ is the number of failures of the ith individual during the time period $[0, t]$. The counting process $\mathbf{N}$ is adapted to the filtration and the sample paths of $N_1, \ldots, N_n$ are right-continuous step functions, zero at time zero, with jumps of size $+1$. No two component processes jump simultaneously. We also assume that N_i has intensity

$$\lambda_i(t) = Y_i(t)\, \alpha(t, Z_i(t)), \tag{2.1}$$

where α is a completely general function, $Y_i(t)$ is a predictable $\{0, 1\}$-valued process indicating that the ith individual is at risk when $Y_i(t) = 1$, and $Z_i(t)$ is a predictable $[0, 1]$-valued covariate process. The processes $M_i(t) = N_i(t) - \int_0^t \lambda_i(s)\, ds$ are local martingales.

Let the cells into which the data are grouped be denoted $C_{rj} = \mathcal{T}_r \times \mathcal{I}_j$, where $\mathcal{T}_1, \ldots, \mathcal{T}_{L_n}$ and $\mathcal{I}_1, \ldots, \mathcal{I}_{J_n}$ are the respective calendar periods (time intervals) and covariate strata. For simplicity, the time intervals are taken to be of equal length $l_n = 1/L_n$ and the covariate strata are taken to have equal width $w_n = 1/J_n$. Grouped data consist of the total number of failures and the total time at risk (exposure) in each cell C_{rj}, given by

$$N_{rj}^{(n)} = \sum_i \int_{\mathcal{T}_r} I\{Z_i(t) \in \mathcal{I}_j\}\, dN_i(t) \quad \text{and} \quad Y_{rj}^{(n)} = \sum_i \int_{\mathcal{T}_r} I\{Z_i(t) \in \mathcal{I}_j\} Y_i(t)\, dt,$$

respectively. All our estimators are based on such data.

Let α_0 be the underlying hazard function to be estimated. To carry out inference for α_0 we need to assume that the support of α_0, denoted $\text{supp}(\alpha_0)$, is known. This assumption is needed to avoid the problem of low exposure on the boundary of the support. In typical survival analysis applications $\text{supp}(\alpha_0)$ can be assumed to be the whole of $[0,1]^2$, but there are simple useful examples where this is not the case; e.g., the illness-death process with duration dependence (see Example 3 of MU) in which $\text{supp}(\alpha_0)$ is the triangle $\{(t,z) \in [0,1]^2 : z < t\}$.

We slightly modify the usual occurrence/exposure rate around the boundary of $\text{supp}(\alpha_0)$ by setting it equal to zero when any part of a cell falls outside $\text{supp}(\alpha_0)$, i.e., define

$$\tilde{\alpha}(t,z) = \frac{N_{rj}^{(n)}}{Y_{rj}^{(n)}} \quad \text{for } (t,z) \in \mathcal{C}_{rj} \subset \text{supp}(\alpha_0),$$

zero otherwise. Our estimators are defined by

$$\tilde{A}(\cdot,z) = \int_0^{\cdot} \tilde{\alpha}(s,z)\,ds \quad \text{and} \quad \tilde{\mathcal{A}}(\cdot,\cdot) = \int_0^{\cdot}\int_0^{\cdot} \tilde{\alpha}(s,x)\,ds\,dx.$$

We assume throughout that α_0 is Lipschitz on its support. The following minor abuse of our notation will be very convenient: for any process (or set) ξ_{rj} indexed by integers r and j, define ξ_{tz} for $(t,z) \in [0,1]^2$ by $\xi_{tz} = \xi_{rj}$ for $(t,z) \in \mathcal{C}_{rj}$. Define

$$\mathcal{D} = \{(t,z) : \mathcal{C}_{tz} \subset \text{supp}(\alpha_0)\},$$

let $\mathcal{D}_z$ denote the z-section of $\mathcal{D}$, and set $Y^{(n)}(t,z) = \sum_i I\{Z_i(t) \in \mathcal{I}_z\} Y_i(t)$.

CONDITION A

(A1) There exists a nonnegative, Lipschitz function $f(\cdot,\cdot)$, which is bounded away from zero on $\text{supp}(\alpha_0)$, such that

$$\int_{\mathcal{D}_z} E\Big(\frac{Y^{(n)}(t,z)}{nw_n} - f(t,z)\Big)^4\,dt = o(l_n^2).$$

(A2) $\text{Leb}\{t \in \mathcal{D}_z : Y^{(n)}(t,z) = 0\} = o_P(l_n)$.

(A3) $\sup_{(t,z)\in\mathcal{D}_z} E\Big(\frac{nw_n}{Y^{(n)}(t,z)}\Big)^4 < \infty$.

(A4) $\text{Leb}(\overline{\mathcal{D}}_z) = O(w_n) + O(l_n)$, where $\overline{\mathcal{D}}_z$ is the set of times t for which $\mathcal{C}_{tz}$ overlaps both $\text{supp}(\alpha_0)$ and its complement.

These conditions are slightly stronger than those required in MU (1990a), but they are still quite mild. In particular, Condition (A4) is satisfied for the illness-death model mentioned above since $\text{Leb}(\overline{\mathcal{D}}_z) \le 2l_n + w_n = O(l_n) + O(w_n)$ in that example.

THEOREM 2.1 *For a fixed $z \in [0,1]$ suppose that Condition A holds. If $nw_n^3 \to 0$, $nw_n l_n^2 \to 0$ and $nw_n \to \infty$, then under $\alpha = \alpha_0$*

$$\sqrt{nw_n}\,(\tilde{A}(\cdot,z) - A(\cdot,z)) \xrightarrow{\mathcal{D}} U(\cdot,z)$$

in $D[0,1]$, where $U(\cdot,z)$ is a continuous Gaussian martingale with zero mean and variance function

$$\mathrm{Var}(U(t,z)) = \int_0^t h(u,z)\,du,$$

where $h = \alpha_0/f$.

In order to derive the asymptotic distribution of $\tilde{A}$ we need to consider a sequence of models of the form (2.1), with each indexed by the sample size and having α piecewise constant over the cells used to group the data, cf. McKeague (1988). At a given sample size α is assumed to be the piecewise constant approximation $\bar{\alpha}_0$ to α_0 determined by the cells C_{ij}. The approximation $\bar{\alpha}_0$ is defined by

$$\bar{\alpha}_0(t,z) = \frac{1}{l_n w_n} \iint_{C_{tz}} \alpha_0(u,x)\,du\,dx \qquad \text{if } C_{tz} \subset \mathrm{supp}(\alpha_0),$$

zero otherwise. The following theorem is an extension to the grouped data setting of MU's (1990b) asymptotic normality result for $\tilde{A}$.

CONDITION B

(B1) $\iint_{\mathcal{D}} E\Big(\frac{Y^{(n)}(t,z)}{n\,w_n} - f(t,z)\Big)^4\,dt\,dz = o(l_n^2 w_n^2)$.

(B2) $\mathrm{Leb}_2\{(t,z) \in \mathcal{D}: Y^{(n)}(t,z) = 0\} = o_P(l_n w_n)$.

(B3) $\sup_{(t,z)\in\mathcal{D}} E\Big(\frac{nw_n}{Y^{(n)}(t,z)}\Big)^4 < \infty$.

THEOREM 2.2 *Suppose that Condition B holds. If $l_n \to 0$, $w_n \to 0$, $nw_n^2 l_n^2 = O(1)$, and $J_n = O(n)$, then the distribution of $\sqrt{n}(\tilde{A} - A)$ under $\alpha = \bar{\alpha}_0$ converges in D_2 to the distribution of the process*

$$\tilde{m}(t,z) = \int_0^z \int_0^t \sqrt{h(u,x)}\,dW(u,x),$$

where $h = \alpha_0/f$ in the interior of $\mathrm{supp}(\alpha_0)$, zero otherwise, and W is a Brownian sheet.

In the following proposition we check that Conditions A and B are satisfied in the i.i.d. case. Let (N_i, Y_i, Z_i), $i = 1,\ldots,n$ be i.i.d. copies of (N, Y, Z). Let $F(t,\cdot)$

be the subdistribution function of the covariate process at time t when $Y(t) = 1$, i.e., $F(t,x) = P(Z(t) \le x, Y(t) = 1)$, $-\infty < x < \infty$.

PROPOSITION 2.1 (i.i.d. case). *Suppose that for each $t \in [0,1]$, $F(t,\cdot)$ is absolutely continuous on the support of $\alpha_0(t,\cdot)$ in $[0,1]$ with density $f(t,\cdot)$ such that $f(\cdot,\cdot)$ is Lipschitz and bounded away from zero on the support of α_0. Suppose that $l_n \to 0$ and $nw_n^3 = O(1)$.*

(i) If $nw_n l_n \to \infty$, then Conditions (A1)-(A3) *hold.*

(ii) If $nw_n^2 l_n \to \infty$, then Condition B *holds.*

When using $\tilde{A}$ in the i.i.d. case it suffices that the interval widths w_n and l_n satisfy $nw_n^3 \to 0$, $nw_n^2 \to \infty$, and $l_n \sim w_n$. When using $\tilde{\mathcal{A}}$ in the i.i.d. case it suffices that the interval widths satisfy $nw_n^3 = O(1)$, $nw_n^{5/2} \to \infty$, and $l_n \sim \sqrt{w_n}$. In particular, this suggests that for $\tilde{\mathcal{A}}$ there should be (asymptotically) more covariate strata than time intervals. Whether such advice should be followed in practice will be considered in the simulation section.

It is possible to give a version of our result for $\tilde{\mathcal{A}}$ under the model $\alpha = \alpha_0$, but we would then need $nw_n^2 \to 0$, which conflicts with the rate in Proposition 2.1 (ii), so the result would not be useful in the i.i.d. case. We are able to get around this difficulty by restricting attention to the sequence of piecewise constant models $\alpha = \bar{\alpha}_0$. We regard this as a very natural approach in the grouped data setting.

3. Fitting the Cox model to grouped data

In the continuous data case the regression coefficient β_0 is estimated by maximizing Cox's partial likelihood function which has logarithm

$$C(\beta) = \sum_i \int_0^1 \beta Z_i(u)\, dN_i(u) - \int_0^1 \log\Big(\sum_i Y_i(u) e^{\beta Z_i(u)}\Big)\, dN^{(n)}(u),$$

where $N^{(n)} = \sum_i N_i$. Pons and Turckheim (1987) estimate β_0 by maximizing a histogram-type Cox's partial likelihood function which has logarithm

$$C_h(\beta) = \sum_r \sum_i \int_{\mathcal{T}_r} \beta Z_i(u)\, dN_i(u) + \sum_r \log\Big(\sum_i \int_{\mathcal{T}_r} e^{\beta Z_i(u)} Y_i(u)\, du\Big) \int_{\mathcal{T}_r} dN^{(n)}(u).$$

However, in the grouped data case neither $C(\beta)$ nor $C_h(\beta)$ is observable. In fact $C_h(\beta)$ is observable with grouped data only when the covariate process Z takes at most finitely many values and is non-time dependent.

For the general grouped data case we need to consider

$$C_g(\beta) = \sum_{r,j} \beta z_j N_{rj}^{(n)} - \sum_r \log\Big(\sum_j Y_{rj}^{(n)} e^{\beta z_j}\Big) N_r^{(n)},$$

where $N_r^{(n)} = \sum_{j=1}^{J_n} N_{rj}^{(n)}$ is the number of failures in the rth calendar period, and z_j is a representative covariate value for the jth stratum. The estimator $\hat{\beta}$ is defined as a solution to $U_g(\beta) = 0$, where U_g is the derivative of C_g. In the piecewise exponential model of Holford (1976), $\hat{\beta}$ is the full maximum likelihood estimator of β_0, and

$$\hat{\lambda}_0(t) = \frac{N_r^{(n)}}{\sum_j Y_{rj}^{(n)} e^{\hat{\beta} z_j}}, \quad \text{for } t \in \mathcal{T}_r,$$

is the maximum likelihood estimator of the baseline hazard function. We also need the grouped data based analogue of Breslow's estimator of the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$ given by

$$\hat{\Lambda}(t) = \int_0^t \hat{\lambda}_0(u)\, du.$$

As in AG, denote

$$S^{(k)}(\beta, t) = \frac{1}{n} \sum_i Z_i^k(t) Y_i(t) e^{\beta Z_i(t)}$$

for $k = 0, 1, 2$, where $0^0 = 1$. The following conditions are assumed to hold throughout this section.

CONDITION C

(C1) (Asymptotic stability). There exists a neighborhood $\mathcal{B}$ of β_0 and function $s^{(k)}$ defined on $\mathcal{B} \times [0, 1]$ such that

$$\sup_{t, \beta \in \mathcal{B}} |S^{(k)}(\beta, t) - s^{(k)}(\beta, t)| \xrightarrow{P} 0 \quad \text{for } k = 0, 1, 2, \tag{C1.1}$$

$$\sup_t E(S^{(k)}(\beta_0, t) - s^{(k)}(\beta_0, t))^2 = O\Big(\frac{1}{n}\Big) \quad \text{for } k = 0, 1. \tag{C1.2}$$

(C2) (Asymptotic regularity). Let $\mathcal{B}$, $s^{(k)}$ be defined as in Condition A and define $v = s^{(2)}/s^{(0)} - (s^{(1)}/s^{(0)})^2$. For all $\beta \in \mathcal{B}$, $t \in [0, 1]$:

$$s^{(1)}(\beta, t) = \frac{\partial}{\partial \beta} s^{(0)}(\beta, t), \quad s^{(2)}(\beta, t) = \frac{\partial^2}{\partial \beta^2} s^{(0)}(\beta, t),$$

$s^{(k)}(\cdot, \cdot)$. $k = 0, 1, 2$, are continuous functions on $\mathcal{B} \times [0, 1]$, $s^{(0)}$ is bounded away from zero on $\mathcal{B} \times [0, 1]$, and

$$\Sigma = \int_0^1 v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt$$

is positive.

We shall also assume that λ_0 is Lipschitz. Our conditions are slightly stronger than the corresponding conditions of AG: we assume a rate of convergence in C1.2, and that $s^{(k)}$ is continuous. Our first result, showing consistency of $\hat{\beta}$, does not need either the full strength of Condition C1 (just C1.1) or the Lipschitz condition. Condition C can be checked in the i.i.d. case, with Z and Y having sample paths in Skorohod space $D[0,1]$, by using similar arguments to Theorem 4.1 of AG.

THEOREM 3.1 *(Consistency of $\hat{\beta}$). If $w_n \to 0$ and $l_n \to 0$, then*

$$\hat{\beta} \xrightarrow{P} \beta_0.$$

THEOREM 3.2 *(Asymptotic normality of $\hat{\beta}$). If $nw_n^2 \to 0$, $nl_n^2 \to 0$ and $nl_n \to \infty$, then*

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma^{-1}).$$

It follows from the proof of Theorem 3.2 that if $w_n \to 0$ and $l_n \to 0$, then $n^{-1} I_g(\hat{\beta})$ is a consistent estimator of Σ, where $I_g(\beta)$ is minus the second derivative of $C_g(\beta)$. The final result in this section gives the limiting distribution of $\hat{\Lambda}$ (cf. AG's Theorem 3.4).

THEOREM 3.3 *If $nw_n^2 \to 0$, $nl_n^2 \to 0$ and $nl_n \to \infty$, then*

$$\sqrt{n}(\hat{\Lambda} - \Lambda_0) \xrightarrow{\mathcal{D}} m_0(\cdot) + m_1(1)\psi(\cdot) \quad \text{in } D[0,1],$$

where

$$\psi(t) = \Sigma^{-1} \int_0^t \frac{s^{(1)}(\beta_0, u)}{s^{(0)}(\beta_0, u)} \lambda_0(u)\, du,$$

and m_0 and m_1 are independent zero mean Gaussian martingales with

$$\langle m_0 \rangle_t = \int_0^t \frac{\lambda_0(u)}{s^{(0)}(\beta_0, u)} du, \quad \langle m_1 \rangle_t = \int_0^t v(\beta_0, u) s^{(0)}(\beta_0, u) \lambda_0(u)\, du.$$

4. Goodness-of-fit test for the Cox model

In this section we consider testing whether an underlying Cox model $\alpha_0(t, z) = \lambda_0(t) \exp\{\beta_0 z\}$ adequately fits the grouped data. Here the support of λ_0 is the whole unit interval. Once again we need to consider a sequence of models $\alpha = \bar{\alpha}_0$ implicitly

indexed by the sample size, where $\bar{\alpha}_0$ is the piecewise constant approximation to α_0 given by

$$\bar{\alpha}_0(t,z) = \bar{\lambda}_0(t)\, e^{\beta_0 z_j} \quad \text{for } z \in \mathcal{I}_j, \tag{4.1}$$

and z_j is a certain covariate value in the jth stratum. Under the Cox model the doubly cumulative hazard function is estimated by

$$\hat{A}(t,z) = \int_0^z \int_0^t \hat{\alpha}(u,x)\, du\, dx,$$

where

$$\hat{\alpha}(u,z) = \hat{\lambda}_0(u)\, e^{\hat{\beta} z_j} \quad \text{for } z \in \mathcal{I}_j.$$

The following result can be used to construct a chi-squared goodness-of-fit test of the Cox model versus the alternative that the Cox model does not hold, based on increments of $\tilde{A} - \hat{A}$, cf. MU (1991).

THEOREM 4.1 *Suppose that Conditions B and C hold. If* $w_n \to 0$, $l_n \to 0$, $nw_n^2 l_n^2 = O(1)$, $J_n = O(n)$ *and* $nl_n \to \infty$, *then the distribution of* $\sqrt{n}(\tilde{A} - \hat{A})$ *under the piecewise constant Cox model* $\alpha = \bar{\alpha}_0$ *converges in* D_2 *to the distribution of the process*

$$m(t,z) = \int_0^z \int_0^t \sqrt{h(u,x)}\, dW(u,x) - b(z) \int_0^1 \int_0^t \frac{\sqrt{g(u,x)}}{s^{(0)}(\beta_0,u)}\, dW(u,x)$$
$$- c(t,z) \int_0^1 \int_0^1 \left\{ x - \frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)} \right\} \sqrt{g(u,x)}\, dW(s,x),$$

where

$$h(u,x) = \frac{\lambda_0(u) e^{\beta_0 x}}{f(u,x)},$$
$$g(u,x) = \lambda_0(u) e^{\beta_0 x} f(u,x),$$
$$b(z) = \int_0^z e^{\beta_0 x}\, dx,$$
$$c(t,z) = \Sigma^{-1} \left\{ \Lambda_0(t) \int_0^z x e^{\beta_0 x} dx - b(z) \int_0^t \frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)} \lambda_0(u)\, du \right\}.$$

When using $\tilde{A} - \hat{A}$ in the i.i.d. case it suffices that w_n and l_n satisfy the same rates as previously given for $\hat{A}$. Note that the rate $nw_n^2 \to 0$ needed for two of our Cox model results (in Section 3) is incompatible with the slower rate needed when using $\tilde{A}$ in the i.i.d. case (see Proposition 2.1 (ii)). For this reason the condition

$nw_n^2 \to 0$ was avoided in Theorem 4.1. Fortunately our Cox model results hold without this condition, provided $\alpha = \bar{\alpha}_0$.

In order to perform the chi-squared test we need to estimate the functions h, g, b and c. This is done by inserting $\hat{\beta}$, $\hat{\lambda}_0$ and estimators of $s^{(k)}(\beta_0, u)$ for $k = 0, 1$ and $f(u, x)$. Estimate $s^{(k)}(\beta_0, u)$ by $S_g^{(k)}(\hat{\beta}, u)$ defined after the proof of Proposition 2.1 in Section 6, and estimate $f(u, x)$ by $Y_{rj}^{(n)}/nw_n l_n$ for $u \in \mathcal{T}_r$ and $x \in \mathcal{I}_j$. Further details on the construction of the chi-squared test can be found in MU (1991).

5. Numerical results

5.1 *Monte Carlo study.* We have carried out a simulation study of the performance of our Cox model estimators and test statistics for various sample sizes, censoring levels, and grouping patterns. The parameters of the underlying Cox model were taken to be $\beta_0 = 1$ and $\lambda_0 \equiv 1$, and the covariate was uniformly distributed on $[0, 1]$. The censoring time was independent of both the failure time and the covariate, and exponentially distributed with parameter γ. The censoring parameter γ was set to 0.75 and 2.5, amounting to 31% and 60% censoring prior to the end of follow-up. In each case the follow-up period was adjusted to give an average of 19% surviving beyond the end of follow-up.

[Insert Table 1 here]

Observed coverage probabilities of asymptotic 95% confidence intervals for β_0 are contained in Table 1. Inspecting Table 1 we find that all the coverage probabilities are close to their nominal value of .95. It appears that variations in sample size and number of cells have little effect. Also, we could find no evidence of bias in $\hat{\beta}$. In view of Holford's (1976) comment that $\hat{\beta}$ is inconsistent unless the grouping becomes finer as the sample size increases, we had expected to eventually obtain poor results at very large sample sizes if the number of cells is kept small. However, this effect only became noticeable for sample sizes above two million.

In order check for asymptotic normality of $\hat{\beta}$ we examined normal plots and histograms of standardized values of $\hat{\beta}$. All of these indicated that $\sqrt{n}(\hat{\beta} - \beta_0)\hat{\Sigma}^{1/2}$ closely follows a standard normal distribution.

We performed the chi-squared goodness-of-fit test of Cox's model using increments of $\tilde{A} - \hat{A}$ over a 2×2 partition of the grouping cells, so that there were 4 degrees of freedom. The results are displayed in Table 2. Very similar results were obtained when using different degrees of freedom, e.g. 4×4. Also, our results are very consistent with those obtained in the continuous data case, see MU (1991).

[Insert Table 2 here]

Inspection of Table 2 indicates that the number of covariate strata J_n has a strong influence on the level accuracy of the test, whereas the number of time

intervals has little effect. For sample sizes less than 1000 we recommend that at most 5 covariate strata be used. For sample sizes between 5000 and 10,000 we recommend that about 10 covariate strata be used. A very large sample size (say $n > 40,000$) would be required to obtain satisfactory results when using 20 covariate strata. These recommendations hold irrespective of the amount of censoring.

We occasionally (in 1–2% of cases) obtained a negative chi-squared statistic at sample size $n = 250$ when using 20 covariate strata. The problem of a negative chi-squared statistic can be caused by small sample size, too many covariate strata, low rate of survival beyond the end of follow-up, or failure of the data to fit the Cox model. Under the Cox model the problem can be avoided by grouping the data so that the total time at risk in each cell is sufficiently large; in the simulation study we found that it was enough to have at least 10% surviving beyond the end of follow-up in each covariate stratum. When the data fail to fit the Cox model the use of a model-robust estimator of the covariance would avoid the problem (cf. Hjort, 1990, p. 1254), but this would be difficult to implement.

5.2. *An example using the Japanese atomic bomb survivors data.* Huffer and McKeague (1991) have studied the application of Aalen's (1980) additive risk model to grouped data on the incidence of cancer mortality among Japanese atomic bomb survivors. It is of interest to examine whether these data can be adequately fitted by the Cox model. The time variable t, taken as time since exposure, is grouped into eight 4-years intervals: 5–9, ..., 33–37 years. The covariate is dose (in units of rads), taken as the midpoint of one of the six dose groups: 0, 1–50, 50–100, 100–200, 200–300, > 300, with dose= 400 for dose > 300. According to our simulation results, this grouping of the data is adequate for the chi-squared test.

Table 3. Results of Cox Model Goodness-of-fit Test for Japanese Atomic Bomb Survivor Data.

	Age at exposure				
	0–9	10–19	20–34	35–49	> 50
Male	reject	reject	negative	reject	reject
Female	negative	reject	negative	negative	reject

We have evaluated the chi-squared statistic separately for males and females in each of 5 different age at exposure groups. There were $4 = 2 \times 2$ degrees of freedom in each test. The results are given in Table 3. The chi-squared statistics indicate extremely strong departures from the Cox model, except in the cases having a negative chi-squared. In fact, as noted above, negative chi-squared values also

suggest a lack of fit with the Cox model when the total time at risk in any cell is sufficiently large (as is the case with these data).

6. Proofs

The following notation will be useful:

$$M_{rj}^{(n)} = \sum_i \int_{\mathcal{T}_r} I\{Z_i(u) \in \mathcal{I}_j\}\, dM_i(u),$$

$$M^{(n)}(t,z) = \sum_i \int_0^t I\{Z_i(u) \in \mathcal{I}_z\}\, dM_i(u),$$

$$\alpha^{(n)}(t,z) = \sum_i I\{Z_i(t) \in \mathcal{I}_z\} Y_i(t) \alpha(t, Z_i(t)),$$

$$\alpha_{tz}^{(n)} = \int_{\mathcal{T}_t} \alpha^{(n)}(u,z)\, du.$$

PROOF OF THEOREM 2.1 Note that

$$\begin{aligned}\sqrt{nw_n}(\tilde{A}(t,z) - A(t,z)) &= X^*(t,z) \\ &+ \sqrt{nw_n} \int_0^t I\{u \in \mathcal{D}_z\} \left(\frac{1}{Y_{uz}^{(n)}} - \frac{1}{nw_n l_n \bar{f}(u,z)} \right) M_{uz}^{(n)}\, du \\ &+ \sqrt{nw_n} \int_0^t \left(I\{u \in \mathcal{D}_z\} \frac{\alpha_{uz}^{(n)}}{Y_{uz}^{(n)}} - \alpha_0(u,z) \right) du,\end{aligned} \tag{6.1}$$

where

$$X(t,z) = \frac{1}{\sqrt{nw_n}} \int_0^t I\{u \in \mathcal{D}_z\} \frac{M^{(n)}(du,z)}{\bar{f}(u,z)},$$

and, given a function ψ defined on [0,1], the piecewise linear approximation ψ^* to ψ determined by the calendar periods $\mathcal{T}_r$ is defined by

$$\psi^*(t) = \psi(t_{r-1}) + \frac{t - t_{r-1}}{l_n} \Big(\psi(t_r) - \psi(t_{r-1}) \Big)$$

for $t \in \mathcal{T}_r = (t_{r-1}, t_r]$. Using an analogous argument to the treatment of the term $X^{(n)}$ in the proof of Theorem 1 of MU (1990a), it can be shown that $X \xrightarrow{\mathcal{D}} U$ in $D[0,1]$. Thus, by Lemma 4.1 of McKeague (1988), we have $X^* \xrightarrow{\mathcal{D}} U$ in $D[0,1]$. The rest of the proof consists in showing that the last two terms on the r.h.s. of (6.1) converge uniformly in probability to zero.

Note that the predictable variation process of M_i is $\langle M_i \rangle_t = \int_0^t \lambda_i(u)du$, so that, using standard martingale theory, $E(M_{uz}^{(n)})^2 \leq O(1)E(Y_{uz}^{(n)})$. Also, from Condition (A1) and the boundedness of f,

$$\int_{\mathcal{D}_z} E(Y_{uz}^{(n)})\, du = o(nw_n l_n^{3/2}) + O(nw_n l_n).$$

It follows that

$$\int_{\mathcal{D}_z} E(M_{uz}^{(n)})^2\, du = O(nw_n l_n). \tag{6.2}$$

The second term on the r.h.s. of (6.1) is of order $O(B_1) + O(B_2)$ uniformly in t, where

$$B_1 = \frac{1}{\sqrt{nw_n}\, l_n} \int_{\mathcal{D}_z} I\{Y_{uz}^{(n)} = 0\}\, |M_{uz}^{(n)}|\, du,$$

$$B_2 = \sqrt{nw_n} \int_{\mathcal{D}_z} \Big| \frac{Y_{uz}^{(n)}}{nw_n l_n} - \bar{f}(u,z) \Big| (Y_{uz}^{(n)})^{-1} |M_{uz}^{(n)}|\, du.$$

Since $Y_{tz}^{(n)} = 0$ implies that $Y^{(n)}(u,z) = 0$ for Lebesgue a.e. $u \in \mathcal{T}_t$, Condition (A2) holds with $Y^{(n)}(t,z)$ replaced by $Y_{tz}^{(n)}$. Thus, by the Cauchy–Schwarz inequality and (6.2), B_1 is of order

$$O\Big(\frac{1}{\sqrt{nw_n}\, l_n}\Big) o_P(\sqrt{l_n}) O_P(\sqrt{nw_n l_n}) = o_P(1).$$

Denote

$$\eta_u = \int_{\mathcal{T}_u} I\{Y^{(n)}(v,z) > 0\}\, dv.$$

Since, for $0 < \delta < 1$,

$$\begin{aligned} P\Big(\inf_{u \in \mathcal{D}_z} \eta_u \leq \delta l_n\Big) &= P\Big(\sup_{u \in \mathcal{D}_z} \int_{\mathcal{T}_u} I\{Y^{(n)}(v,z) = 0\}\, dv > l_n(1-\delta)\Big) \\ &\leq P(\mathrm{Leb}\{u \in \mathcal{D}_z : Y^{(n)}(u,z) = 0\} > l_n(1-\delta)) \to 0 \end{aligned}$$

by Condition (A2), we see that η_u^{-1} is of order $O_P(l_n^{-1})$ uniformly in $u \in \mathcal{D}_z$. The Cauchy–Schwarz inequality gives

$$\eta_u^2 \leq Y_{uz}^{(n)} \int_{\mathcal{T}_u} (Y^{(n)}(v,z))^{-1}\, dv,$$

which yields an upper bound on $(Y_{uz}^{(n)})^{-1}$, so that

$$B_2 \leq O_P\Big(\frac{\sqrt{nw_n}}{l_n^2}\Big) \int_{\mathcal{D}_z} \Big| \frac{Y_{uz}^{(n)}}{nw_n l_n} - \bar{f}(u,z) \Big| \Big(\int_{\mathcal{T}_u} \frac{1}{Y^{(n)}(v,z)}\, dv \Big) |M_{uz}^{(n)}|\, du. \tag{6.3}$$

By Condition (A1),

$$E\int_{\mathcal{D}_z}\Big(\frac{Y_{uz}^{(n)}}{nw_nl_n}-\bar{f}(u,z)\Big)^4\,du=o(l_n^2).$$

Also, by (A3),

$$E\int_{\mathcal{D}_z}\Big[\int_{\mathcal{T}_u}(Y^{(n)}(v,z))^{-1}\,dv\Big]^4\,du=O\Big(\frac{l_n}{nw_n}\Big)^4.$$

Thus, using (6.2) and the Cauchy–Schwarz inequality twice, the expectation of the integrand in (6.3) is of order

$$\Big\{\Big[o(l_n^2)\Big]^{\frac{1}{2}}\cdot\Big[O\Big(\frac{l_n}{nw_n}\Big)^4\Big]^{\frac{1}{2}}\Big\}^{\frac{1}{2}}\cdot\Big\{O(nw_nl_n)\Big\}^{\frac{1}{2}}=o\Big(\frac{l_n^2}{\sqrt{nw_n}}\Big).$$

It follows that B_2 is of order $o_P(1)$.

For $u\in(\mathcal{D}_z\cup\overline{\mathcal{D}}_z)^c$, $\alpha_0(u,z)=0$. By the Lipschitz assumption on α_0, the last term on the r.h.s. of (6.1) is bounded uniformly in t by

$$\begin{aligned}&\sqrt{nw_n}\int_{\mathcal{D}_z}\Big|\frac{[\alpha_0(u,z)+O(l_n)+O(w_n)]Y_{uz}^{(n)}}{Y_{uz}^{(n)}}-\alpha_0(u,z)\Big|du+\sqrt{nw_n}\int_{\overline{\mathcal{D}}_z}|\alpha_0(u,z)|du\\&\le\sqrt{nw_n}\{O(l_n)+O(w_n)+O(1)\mathrm{Leb}\{u\in\mathcal{D}_z:Y_{uz}^{(n)}=0\}\}+O(\sqrt{nw_n})\mathrm{Leb}\{\overline{\mathcal{D}}_z\}\\&\xrightarrow{P}0\end{aligned}$$

by (A2), (A4), $nw_n^3\to0$ and $nw_nl_n^2\to0$. This completes the proof. □

PROOF OF THEOREM 2.2 By routine calculation

$$\begin{aligned}\sqrt{n}(\tilde{\mathcal{A}}-\mathcal{A})(t,z)&=X^\dagger(t,z)\\&+\sqrt{n}\int_0^z\int_0^t I\{u\in\mathcal{D}_x\}\Big(\frac{1}{Y_{ux}^{(n)}}-\frac{1}{nw_nl_n\bar{f}(u,x)}\Big)M_{ux}^{(n)}\,du\,dx\\&+\sqrt{n}\int_0^z\int_0^t I\{u\in\mathcal{D}_x\}\Big(\frac{\alpha_{ux}^{(n)}}{Y_{ux}^{(n)}}-\bar{\alpha}_0(u,x)\Big)\,du\,dx,\end{aligned}\tag{6.4}$$

where

$$X(t,z)=\frac{1}{\sqrt{n}}\sum_{j=1}^{[zJ_n]}\int_0^t I\{u\in\mathcal{D}_x\}\frac{M^{(n)}(du,z_j)}{\bar{f}(u,z_j)},$$

and given a function ψ on $[0,1]^2$, $\psi^\dagger$ is defined to be the piecewise bilinear approximation to ψ determined by the cells C_{rj}, obtained by extending the definition of ψ^* in the obvious way.

To complete the proof it suffices to show that the last two terms of (6.4) converge uniformly in probability to 0, and $X \xrightarrow{\mathcal{D}} \tilde{m}$, where we are using Lemma 4.1 of McKeague (1988) again. As in the proof of Theorem 3.1 of MU (1990b), we need to show that $\{X,\ n \geq 1\}$ is tight in D_2 and its finite dimensional distributions converge weakly to those of $\tilde{m}$. Tightness can be shown using $J_n = O(n)$, see Lemmas 2 and 3 of MU (1990b). Convergence of the finite dimensional distributions can be shown using Condition (B1) to obtain convergence of the predictable variation processes of increments of X (cf. Lemma 4 of MU (1990b)), and verifying the Lindeberg condition which is similar to (3.2) of MU (1990b).

The second term on the r.h.s. of (6.4) is uniformly of order $O(G_1) + O(G_2)$, where

$$G_1 = \frac{1}{\sqrt{n} w_n l_n} \iint_{\mathcal{D}} I\{Y_{ux}^{(n)} = 0\} \, |M_{ux}^{(n)}| \, du \, dx,$$

$$G_2 = \sqrt{n} \iint_{\mathcal{D}} \Big| \frac{Y_{ux}^{(n)}}{n w_n l_n} - \bar{f}(u,x) \Big| (Y_{ux}^{(n)})^{-1} |M_{ux}^{(n)}| \, du \, dx.$$

As in dealing with B_1 in the proof of Theorem 2.1, the Condition (B2) holds with $Y^{(n)}(t,z)$ replaced by $Y_{tz}^{(n)}$. By Condition (B1) and boundedness of f, we have that $\int\int_{\mathcal{D}} E(M_{ux}^{(n)})^2 du dx = O(n w_n l_n)$. Thus, G_1 is of order

$$\frac{1}{\sqrt{n} w_n l_n} o_P\Big(\sqrt{w_n l_n}\Big) O_P\Big(\sqrt{n w_n l_n}\Big) = o_P(1).$$

Also, as in dealing with B_2 in the proof of Theorem 2.1, define

$$\eta_{ux} = \int_{\mathcal{T}_u} I\{Y^{(n)}(v,x) > 0\} \, dv.$$

Since, for $0 < \delta < 1$,

$$\begin{aligned} P\Big(\inf_{(u,x) \in \mathcal{D}} \eta_{ux} \leq \delta l_n \Big) &= P\Big(\sup_{(u,x) \in \mathcal{D}} \int_{\mathcal{T}_u} I\{Y^{(n)}(v,x) = 0\} \, dv > l_n(1-\delta) \Big) \\ &\leq P\Big(\iint_{\mathcal{D}} I\{Y^{(n)}(u,x) = 0\} \, du dx > w_n l_n (1-\delta) \Big) \\ &\leq P(\mathrm{Leb}\{(u,x) \in \mathcal{D} : Y^{(n)}(u,x) = 0\} > w_n l_n (1-\delta)) \to 0 \end{aligned}$$

by Condition (B2), we have that η_{ux}^{-1} is of order $O_P(l_n^{-1})$ uniformly in $(u,x) \in \mathcal{D}$. So that

$$G_2 \leq O_P\Big(\frac{\sqrt{n}}{l_n^2}\Big) \iint_{\mathcal{D}} \Big| \frac{Y_{ux}^{(n)}}{n w_n l_n} - \bar{f}(u,x) \Big| \Big(\int_{\mathcal{T}_u} \frac{1}{Y^{(n)}(v,x)} \, dv \Big) |M_{ux}^{(n)}| \, du \, dx. \qquad (6.5)$$

By (B1), (B3) and Cauchy-Schwarz equality, the expectation of the integrand in (6.5) is of order

$$\left\{\left[o_P(l_n^2 w_n^2)\right]^{\frac{1}{2}} \cdot \left[O_P\left(\frac{l_n}{nw_n}\right)^4\right]^{\frac{1}{2}}\right\}^{\frac{1}{2}} \cdot \left\{O_P(nw_n l_n)\right\}^{\frac{1}{2}} = o_P\left(\frac{l_n^2}{\sqrt{n}}\right).$$

It follows that G_2 is of order $o_P(1)$. Thus the second term converges uniformly in probability to zero.

Consider the last term of (6.4). For the piecewise constant model, $\alpha_{ux}^{(n)} = \bar{\alpha}_0(u,x)Y_{ux}^{(n)}$. Thus, the last term of (6.4) is of order

$$O(\sqrt{n})\text{Leb}_2\{(u,x): Y_{ux}^{(n)} = 0\} \xrightarrow{P} 0$$

by Condition (B2) and $nw_n^2 l_n^2 = O(1)$. This completes the proof. □

PROOF OF PROPOSITION 2.1 We only consider part (i); the proof of part (ii) is almost identical. Under the conditions of the proposition, $Y^{(n)}(t,z)$ has a binomial distribution with parameters n and $\int_{\mathcal{I}_z} f(t,x)\,dx$, and there exist positive constants b and c such that $bw_n \le \int_{\mathcal{I}_z} f(t,x)\,dx \le cw_n$, for each $(t,z) \in \text{supp}(\alpha_0)$. Since the fourth central moment of a binomial (n,p) r.v. is of order $O((np)^2)$, and using the Lipschitz assumption on f, as well as $nw_n^3 = O(1)$ and $nw_n l_n \to \infty$, we have

$$E\left(\frac{Y^{(n)}(t,z)}{nw_n} - f(t,z)\right)^4 = O\left(\frac{(ncw_n)^2}{(nw_n)^4}\right) + O(w_n^4) = o(l_n^2)$$

uniformly over $(t,z) \in \text{supp}(\alpha_0)$. Condition (A1) follows using the dominated convergence theorem. Next,

$$E[\text{Leb}\{t \in \mathcal{D}_z: Y^{(n)}(t,z) = 0\}] \le (1 - bw_n)^n \le e^{-bnw_n} = o(l_n)$$

by $nw_n l_n \to \infty$, giving Condition (A2). Condition (A3) holds by Lemma 2(i) of MU (1990a). □

It is useful to define the grouped data version of $S^{(k)}$ given by

$$S_g^{(k)}(\beta,t) = \frac{1}{nl_n}\sum_j z_j^k Y_{rj}^{(n)} e^{\beta z_j} \quad \text{for } t \in \mathcal{T}_r.$$

Note that, by Conditions (C1.1) and continuity of $s^{(k)}$ in (C2), if $w_n \to 0$ and $l_n \to 0$, then $S_g^{(k)}$ satisfies Condition (C1.1) in place of $S^{(k)}$.

PROOF OF THEOREM 3.1 Define $X(\beta) = n^{-1}(C(\beta) - C(\beta_0))$, and $X_g(\beta)$ analogously for the grouped data case. Consider the difference between $X(\beta)$ and $X_g(\beta)$, for some fixed β. Routine manipulation gives

$$\begin{aligned}\left|X_g(\beta) - X(\beta)\right| \leq & \frac{1}{n}\left|\sum_{r,j,i}\int_{\mathcal{T}_r}(\beta-\beta_0)(z_j - Z_i(u))I\{Z_i(u)\in\mathcal{I}_j\}dN_i(u)\right| \\ & + \frac{1}{n}\int_0^1 \left|\log\left(\frac{S_g^{(0)}(\beta,u)}{S_g^{(0)}(\beta_0,u)}\right) - \log\left(\frac{S^{(0)}(\beta,u)}{S^{(0)}(\beta_0,u)}\right)\right| dN^{(n)}(u).\end{aligned}$$

The first term on the r.h.s. is bounded above by

$$|\beta-\beta_0|\sup_{i,j,u}\left|(z_j - Z_i(u))I\{Z_i(u)\in\mathcal{I}_j\}\right| \cdot \frac{1}{n}N^{(n)}(1) \xrightarrow{P} 0,$$

since the width of $\mathcal{I}_j$ is $w_n \to 0$ and $n^{-1}N^{(n)}(1) = O_P(1)$, see AG (p. 1108). Similarly, the second term tends in probability to zero by continuity of log, the remark preceding the proof, and the assumption that $s^{(0)}$ is bounded away from zero. Thus $|X_g(\beta) - X(\beta)| \xrightarrow{P} 0$. The result now follows using the argument in Section 2.3 of AG. □

PROOF OF THEOREM 3.2 Using a Taylor expansion of $U_g(\beta)$ around β_0, and inspecting the proof in the continuous data case (AG, p. 1106), we reduce to showing that

$$\frac{1}{\sqrt{n}}\left|U_g(\beta_0) - U(\beta_0)\right| \xrightarrow{P} 0 \quad \text{and} \quad \frac{1}{n}\sup_{\beta\in\mathcal{B}}\left|I_g(\beta) - I(\beta)\right| \xrightarrow{P} 0,$$

where $U(\beta)$ is the derivative of $C(\beta)$, and $I(\beta)$ is minus the second derivative of $C(\beta)$. Now, since

$$\begin{aligned}U_g(\beta_0) &= \sum_{r,j} z_j N_{rj}^{(n)} - \sum_r \left(\frac{\sum_j z_j e^{\beta_0 z_j} Y_{rj}^{(n)}}{\sum_j e^{\beta_0 z_j} Y_{rj}^{(n)}}\right) N_r^{(n)}, \\ U(\beta_0) &= \sum_i \int_0^1 Z_i(u)\, dM_i(u) - \int_0^1 \frac{S^{(1)}(\beta_0,u)}{S^{(0)}(\beta_0,u)}\, dM^{(n)}(u),\end{aligned}$$

where $M^{(n)} = \sum_i M_i$, we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\left|U_g(\beta_0) - U(\beta_0)\right| \\ &\qquad \leq \frac{1}{\sqrt{n}}\left|\sum_{j,i}\int_0^1 (z_j - Z_i(u))I\{Z_i(u)\in\mathcal{I}_j\}\, dM_i(u)\right| \qquad (6.6)\end{aligned}$$

$$+\frac{1}{\sqrt{n}}\Big|\int_0^1\left(\frac{S^{(1)}(\beta_0,u)}{S^{(0)}(\beta_0,u)}-\frac{S_g^{(1)}(\beta_0,u)}{S_g^{(0)}(\beta_0,u)}\right)dM^{(n)}(u)\Big| \tag{6.7}$$

$$+\frac{1}{\sqrt{n}}\Big|\int_0^1\Bigg(\sum_{j,i}z_j e^{\beta_0 Z_i(u)}I\{Z_i(u)\in\mathcal{I}_j\}Y_i(u)$$

$$-n\frac{S_g^{(1)}(\beta_0,u)}{S_g^{(0)}(\beta_0,u)}S^{(0)}(\beta_0,u)\Bigg)\lambda_0(u)\,du\Big| \tag{6.8}$$

The integrands in (6.6) are predictable and bounded in modulus by w_n, so that, using standard martingale theory, (6.6) is of order $O_P(w_n)$.

Next, (6.7) is bounded by

$$\frac{1}{\sqrt{n}}\Big|\int_0^1\left(\frac{S^{(1)}(\beta_0,u)}{S^{(0)}(\beta_0,u)}-\frac{\bar{s}^{(1)}(\beta_0,u)}{\bar{s}^{(0)}(\beta_0,u)}\right)dM^{(n)}(u)\Big|$$

$$+\frac{1}{\sqrt{n}}\sum_r\Big|\frac{\bar{s}^{(1)}(\beta_0,t_r)}{\bar{s}^{(0)}(\beta_0,t_r)}-\frac{S_g^{(1)}(\beta_0,t_r)}{S_g^{(0)}(\beta_0,t_r)}\Big||M_r^{(n)}|,$$

where $M_r^{(n)}$ is the increment of $M^{(n)}$ over $\mathcal{T}_r$ and t_r is an arbitrary point in $\mathcal{T}_r$. The integrand of the first term is predictable and of order $o_P(1)$ uniformly in u, by Conditions (C1.1) and (C2), so, as in dealing with (6.6), the first term above converges in probability to zero. To deal with the second term, approximate $S_g^{(1)}/S_g^{(0)}$ by $\bar{S}^{(1)}/\bar{S}^{(0)}$, with an error of order $O_P(w_n)$ uniformly in t. This leads to the upper bound

$$\frac{1}{\sqrt{n}}O_P(1)\sum_{k=0}^{1}\sum_{r=1}^{L_n}|\bar{s}^{(k)}(\beta_0,t_r)-\bar{S}^{(k)}(\beta_0,t_r)||M_r^{(n)}| \tag{6.9}$$

$$+O_P(w_n)\frac{1}{\sqrt{n}}\sum_r|M_r^{(n)}|. \tag{6.10}$$

Note that the second moment of $M_r^{(n)}$ is of order $O(nl_n)$ uniformly in r, so, by the Cauchy–Schwarz inequality and Condition (C1.2), we have

$$E[|\bar{s}^{(k)}(\beta_0,t_r)-\bar{S}^{(k)}(\beta_0,t_r)||M_r^{(n)}|]\le O(n^{-1})^{1/2}O(nl_n)^{1/2}=O(l_n)^{1/2},$$

uniformly in r, for $k=0,1$. Thus (6.9) is of order

$$n^{-1/2}O_P(1)L_nO(l_n)^{1/2}=O_P(nl_n)^{-1/2}\xrightarrow{P}0$$

since $nl_n\to\infty$. The term (6.10) is of order

$$O_P(w_n)n^{-1/2}L_nO(nl_n)^{1/2}=O_P(w_n^2/l_n)^{1/2}\xrightarrow{P}0,$$

since $nw_n^2 \to 0$ and $nl_n \to \infty$. We have shown that (6.7) converges in probability to zero.

Now consider the lase term (6.8). Approximate the first term of the integrand by $nS^{(1)}$, with an error of order $O(nw_n)$ uniformly in u. Then, also approximating $S_g^{(1)}/S_g^{(0)}$ by $\bar{S}^{(1)}/\bar{S}^{(0)}$, we obtain an upper bound of order

$$O_P(\sqrt{n}w_n) + \sqrt{n}\Big| \int_0^1 \Big(S^{(1)}(\beta_0, u) - \frac{\bar{S}^{(1)}(\beta_0, u)}{\bar{S}^{(0)}(\beta_0, u)} S^{(0)}(\beta_0, u)\Big) \lambda_0(u)\, du\Big|.$$

By the Lipschitz condition on λ_0, the second term is of order $O_P(\sqrt{n}l_n)$. Since $nw_n^2 \to 0$ and $nl_n^2 \to 0$, it follows that (6.8) tends in probability to zero.

Finally, consider the difference between $I_g(\beta)$ and $I(\beta)$. Note that

$$I(\beta) = \int_0^1 \Big\{ \frac{S^{(2)}(\beta, u)}{S^{(0)}(\beta, u)} - \Big(\frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)}\Big)^2 \Big\}\, dN^{(n)}(u),$$

and $I_g(\beta)$ is obtained by replacing $S^{(k)}$ by $S_g^{(k)}$ throughout this expression. By the remark preceding the proof of Theorem 2.1, the difference between the integrands in $I_g(\beta)$ and $I(\beta)$ tends uniformly in probability to 0. Since $N^{(n)}(1) = O_P(n)$, it follows that $n^{-1} \sup_\beta |I_g(\beta) - I(\beta)| \xrightarrow{P} 0$, completing the proof. □

PROOF OF THEOREM 3.3 We shall use the notation $N^g(u) = N_r^{(n)}$ for $u \in \mathcal{T}_r$. Note that

$$\begin{aligned} \sqrt{n}(\hat{\Lambda}(t) - \Lambda_0(t)) =& \frac{1}{\sqrt{n}l_n} \int_0^t \Big\{ \frac{1}{S_g^{(0)}(\hat{\beta}, u)} - \frac{1}{\bar{S}^{(0)}(\hat{\beta}, u)} \Big\} N^g(u)\, du \\ &+ X^*(t) + \sqrt{n}\Big\{ \int_0^t \frac{N^g(u)}{nl_n \bar{S}^{(0)}(\beta_0, u)}\, du - \Lambda_0(t) \Big\}, \end{aligned} \tag{6.11}$$

where

$$X(t) = \frac{1}{\sqrt{n}} \int_0^t \Big(\frac{1}{\bar{S}^{(0)}(\hat{\beta}, u)} - \frac{1}{\bar{S}^{(0)}(\beta_0, u)} \Big)\, dN^{(n)}(u).$$

Consider the first term on the right hand side of the decomposition (6.11). We may approximate $S_g^{(0)}(\hat{\beta}, u)$ by $\bar{S}^{(0)}(\hat{\beta}, u)$ with an error of order $O(w_n)$ uniformly in u. The second moment of $M_r^{(n)}$ is of order of $O(nl_n)$ uniformly in r, so that $N_r^{(n)}$ is of order of $O_P(nl_n)$ uniformly in r by $nl_n \to \infty$. Thus the first term is of order

$$O_P(w_n) \frac{1}{\sqrt{n}l_n} O_P(nl_n) = O_P(\sqrt{n}w_n) \xrightarrow{P} 0$$

uniformly in t and β.

As in AG, denote

$$H(\beta,t) = -\frac{1}{n}\int_0^t \frac{S^{(1)}(\beta,u)}{\{S^{(0)}(\beta,u)\}^2}\, dN^{(n)}(u),$$

and let H_g be the grouped version of H obtained by replacing $S^{(k)}$ by $\bar{S}^{(k)}$, $k = 0, 1$. Taylor's expansion gives

$$X(t) = H_g(\beta_1, t) \cdot \sqrt{n}(\hat{\beta} - \beta_0),$$

where β_1 lies between $\hat{\beta}$ and β_0. From the proof of Theorem 3.4 of AG (p.1109), $H(\beta_1, t)$ converges in probability to $-\Sigma \cdot \psi(t)$ uniformly in t, for any β_1 such that $\beta_1 \xrightarrow{P} \beta_0$. The difference between $H_g(\beta, t)$ and $H(\beta, t)$ is of order $o_P(1)$ uniformly in β and t. Thus, $H_g(\beta_1, t)$ also converges in probability to $-\Sigma \cdot \psi(t)$ and it follows from the proof of Theorem 3.2 and the proof of Theorem 3.2 of AG (p. 1106) that $X \xrightarrow{\mathcal{D}} m_1(1)\psi$. We have used the conditions $nw_n^2 \to 0$, $nl_n^2 \to 0$ and $nl_n \to \infty$ in order to be able to appeal to Theorem 3.2.

Since λ_0 is Lipschitz and $nl_n^2 \to 0$, as in dealing with (6.7) in the proof of Theorem 3.2, the last term on the right hand side of the decomposition of (6.11) differs from the piecewise linear approximation M_0^* to

$$M_0(t) = \frac{1}{\sqrt{n}}\int_0^t \frac{dM^{(n)}(u)}{s^{(0)}(\beta_0, u)}$$

by at most $o_P(1)$, uniformly in t. From the proof of Theorem 3.4 of AG, $M_0 \xrightarrow{\mathcal{D}} m_0$ and, also using the proof of Theorem 3.2, M_0 and X are seen to be asymptotically independent. Thus (M_0, X) converges in distribution to $(m_0, m_1(1)\psi)$, and, by applying a slightly extended version of Lemma 4.1 of McKeague (1988), so does (M_0^*, X^*). This completes the proof. □

PROOF OF THEOREM 4.1 The proof is almost identical to the proof of Theorem 4.1 of MU (1991), so we only sketch the required modifications. In (6.8) of that paper, $\tilde{M}^{\dagger}$ replaces $\tilde{M}$, M_0^* replaces $\widehat{M}_0$, and we appeal to Lemma 4.1 of McKeague (1988) to show that these substitutions make no difference asymptotically. We also need to show that

$$s^{(k)}(\beta_0, u) = \int_0^1 x^k e^{\beta_0 x} f(u, x)\, dx \quad \text{for } k = 0, 1, 2, \tag{6.12}$$

as is immediate in the i.i.d. case. Since $x \mapsto x^k e^{\beta_0 x}$ is Lipschitz on bounded intervals,

$$\int_0^1 \left| S^{(k)}(\beta_0, u) - \int_0^1 x^k e^{\beta_0 x} f(u, x)\, dx \right| du = o_P(\sqrt{w_n l_n}) + O_P(w_n)$$

by Condition (B1). Thus the representation (6.12) follows by the triangle inequality, Condition (C1.1), and continuity of $s^{(k)}$.

Finally, note that under the piecewise constant model the remainder terms (6.8), (6.10) and the first term in (6.11) are absent, so Theorems 3.2 and 3.3 hold without the condition $nw_n^2 \to 0$. □

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Tech. Report, Dept. of Statistics, University of California, Berkeley.

Borgan, Ø. and Ramlau-Hansen, H. (1985). Demographic incidence rates and estimation of intensities with incomplete information. *Ann. Statist.* **13** 564–582.

Breslow, N. E. (1986). Cohort analysis in epidemiology, in A. C. Atkinson and S. E. Fienberg, eds., *A Celebration of Statistics: the ISI Centenary Volume.* Springer-Verlag, New York, 109–143.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B.* **34** 187–220.

Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *Ann. Statist.* **10** 101–113.

Hjort, N. L. (1990). Goodness of fit tests in models for life history data based on cumulative hazard rates. *Ann. Statist.* **18** 1221–1258.

Hoem, J. M. (1987). Statistical analysis of a multiplicative model and its application to the standardization of vital rates: a review. *Int. Statist. Rev.* **55** 119–152.

Holford, T. R. (1976). Life tables with concomitant information. *Biometrics* **32** 587–597.

Huffer, F. W. and McKeague, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J. Amer. Statist. Assoc.* **86** 38–53.

Kalbfleisch J. D. and Prentice R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60** 267–278.

Karr, A. F. (1987). Maximum likelihood estimation in the multiplicative intensity model via sieves. *Ann. Statist.* **15** 473–490.

Laird, N. and Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *J. Amer. Statist. Assoc.* **76** 231–240.

Marsaglia, G., Zaman, A. and Tsang, W. W. (1990). Toward a universal random number generator. *Statistics and Probability Letters* **8** 35–39.

McKeague, I. W. (1988). A counting process approach to the regression analysis of grouped survival data. *Stoch. Process. Appl.* **28** 221–239.

McKeague, I. W. and Utikal, K. J. (1990a). Inference for a nonlinear counting process regression model. *Annals of Statistics* **18** 1172–1187.

McKeague, I. W. and Utikal, K. J. (1990b). Identifying nonlinear covariate effects in semimartingale regression models. *Probability Theory and Related Fields* **87** 1–25.

McKeague, I. W. and Utikal, K. J. (1991). Goodness-of-fit tests for additive hazards and proportional hazards models. Submitted to *Scandinavian Journal of Statistics.*

Pons O. and Turckheim, E. de (1987). Estimation in Cox's periodic model with a histogram-type estimator for the underlying intensity. *Scand. J. Statist.* **14** 329–345.

Prentice R. L. and Gloeckler L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* **34** 57–67.

Selmer, R. (1990). A comparison of Poisson regression models fitted to multiway summary tables and Cox's survival model using data from a blood pressure screening in the city of Bergen, Norway. *Statistics in Medicine* **9** 1157–1165.

Table 1. Observed Coverage Probabilities of Asymptotic 95% Confidence Intervals for β_0.

Cens	L_n	J_n	Sample Size					
			250	500	1000	5000	10000	50000
31%	5	5	.9490	.9494	.9492	.9496	.9506	.9438
	10	5	.9490	.9488	.9496	.9496	.9504	.9456
	20	5	.9488	.9490	.9498	.9494	.9502	.9450
	5	10	.9510	.9478	.9466	.9508	.9482	.9480
	10	10	.9510	.9478	.9476	.9508	.9490	.9484
	5	20	.9492	.9464	.9494	.9504	.9508	.9478
	20	20	.9488	.9458	.9494	.9502	.9512	.9474
60%	5	5	.9516	.9490	.9512	.9428	.9460	.9526
	10	5	.9524	.9494	.9514	.9428	.9452	.9528
	20	5	.9520	.9492	.9516	.9428	.9452	.9526
	5	10	.9518	.9460	.9484	.9454	.9476	.9530
	10	5	.9528	.9466	.9480	.9454	.9480	.9530
	5	20	.9536	.9466	.9502	.9460	.9468	.9522
	20	20	.9524	.9462	.9506	.9452	.9470	.9526

Table 2. Observed Probabilites of Rejecting the Cox Model at Asymptotic Level 5%.

Cens	L_n	J_n	Sample Size					
			250	500	1000	5000	10000	50000
31%	5	5	.0806	.0556	.0608	.0568	.0516	.0472
	10	5	.0732	.0596	.0576	.0576	.0520	.0472
	20	5	.0750	.0584	.0594	.0570	.0484	.0480
	5	10	.2160	.1480	.1180	.0704	.0608	.0510
	10	10	.2298	.1542	.1146	.0800	.0658	.0564
	5	20	.6534	.4290	.3236	.1744	.1276	.0668
	20	20	.7364	.4702	.3326	.1836	.1368	.0772
60%	5	5	.0702	.0580	.0634	.0592	.0526	.0520
	10	5	.0644	.0580	.0548	.0576	.0504	.0494
	20	5	.0644	.0582	.0576	.0600	.0490	.0478
	5	10	.1592	.1176	.1006	.0796	.0682	.0572
	10	10	.1600	.1114	.0964	.0800	.0668	.0578
	5	20	.4662	.2830	.2196	.1576	.1308	.0798
	20	20	.5306	.3056	.2222	.1622	.1348	.0888

NOTES: The data were generated using the uniform random number generator of Marsaglia, Zaman and Tsang (1990). The number of samples in each run was 5000.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

MASTER COPY - FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) FSU Technical Report No. M-840	5. MONITORING ORGANIZATION REPORT NUMBER(S) USARO Technical Report No. D-117
6a. NAME OF PERFORMING ORGANIZATION Florida State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics Tallahassee, FL 32306-3033	7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U. S. Army Research Office 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) On the analysis of grouped survival data using cumulative occurrence/exposure rates	
12. PERSONAL AUTHOR(S) Ian W. McKeague and Mei-Jie Zhang	
13a. TYPE OF REPORT Technical 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) March 1991 15. PAGE COUNT 25
16. SUPPLEMENTARY NOTATION The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.	
17. COSATI CODES FIELD / GROUP / SUB-GROUP	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Counting processes, conditional hazard function, grouped survival data, martingale central limit theorem, proportional hazards model, nonlinear regression.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

See Back

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

Abstract

One may estimate a conditional hazard function from grouped (and possibly censored) survival data by the time and covariate specific occurrence/exposure rate. Asymptotic results for cumulative versions of this estimator are developed, utilizing the general framework of counting processes. In particular, a g ıped data based goodness-of-fit test for Cox's proportional hazard model is given. Various constraints on the asymptotic behavior of the widths of the calendar periods and covariate strata employed in grouping the data are needed to prove the results. Actual performance of the estimators and test statistics is evaluated by Monte Carlo methods.